ON A GENERAL CONDITION FOR NULL ROBUSTNESS

by

Morris L. Eaton[1]
University of Minnesota

Takeaki Kariya[2]
University of Pittsburgh
and
Hitotsubashi University

Technical Report No. 388

June 1981

AMS 1970 Subject Classification: Primary - 62E10; Secondary -62H05, 62H10.

Key Words and Phrases: Null robustness, Invariance, MANOVA, Canonical correlations, Tests of independence, Multivariate normal distribution theory, Complex multivariate normal.

[1]This research was supported in part by a National Science Foundation Grant MCS 77-25112.

[2]This research was supported in part by an Andrew Mellon Fellowship.

ABSTRACT

This paper gives sufficient conditions that certain statistics have a common distribution under a wide class of underlying distributions. Invariance methods are the primary technical tool in establishing the theoretical results. These results are applied to MANOVA problems, problems involving canonical correlations, and certain statistics associated with the complex normal distribution.

1. INTRODUCTION

Using a geometric argument, Fisher (1925) showed that Student's one sample t-statistic has the same null distribution under normality as under the assumption of spherical symmetry (see Efron (1969) for a discussion and some related topics). This fact about the t-statistic is due to two things:

(i) the t-statistic is scale invariant

(ii) the uniform distribution on S_{n-1} (the sphere of radius 1 in R^n) is the unique spherically symmetric distribution on S_{n-1}.

Basically, we have used this observation to formulate a general method for proving similar results about other statistics of interest. Some additional properties of the t-test under spherical symmetry are established in Kariya and Eaton (1977). Recently, multivariate analogues of some of these results have been developed by Dawid (1977), Fraser and Ng (1980), Kariya (1981a), and Kariya (1981b).

Let $(\mathcal{X},\mathcal{B})$ be a measurable space and suppose X takes values in $\mathcal{X}$. We write $\mathcal{L}(X) = P$ to mean that the distribution of X is P. If t(X) is any statistic, $\mathcal{L}(t(X)|P)$ denotes the distribution of t(X) when $\mathcal{L}(X) = P$. Now, suppose that the distribution of t(X) is known when $\mathcal{L}(X) = P_0$ and set

$$\mathcal{P} = \{P \,|\, \mathcal{L}(t(X)|P) = \mathcal{L}(t(X)|P_0)\}. \tag{1.1}$$

It would be of interest to supply some useful sufficient conditions which imply that $P \in \mathcal{P}$. Using invariance assumptions, this is what is done in Theorems 2.1 and 2.2. An example from multivariate analysis will help expose one of the underlying ideas.

Example 1.1: Take $\mathcal{X}$ to be the set of all $n \times p$ real matrices of rank p (so $n \geq p$) and suppose $X \in \mathcal{X}$. A number of important statistics which arise in testing problems in MANOVA (see Section 4) can be written as functions of

$$T = t(X) = X(X'X)^{-1}X'$$

which is a random orthogonal projection of rank p. The distributions of these functions of $t(X)$ can often be computed when the elements of X are i.i.d. $N(0,1)$ -- let this be P_0. We now assert that if $\mathcal{L}(X) = \mathcal{L}(\Gamma X)$ for each $\Gamma \in \mathcal{O}_n$ (the orthogonal group), then $\mathcal{L}(X) \in \mathcal{P}$ which is defined by (1.1). To see this, first observe that $t(\Gamma X) = \Gamma t(X)\Gamma'$ so

$$(*) \qquad \mathcal{L}(T) = \mathcal{L}(\Gamma T \Gamma'), \qquad \Gamma \in \mathcal{O}_n$$

when $\mathcal{L}(X) = \mathcal{L}(\Gamma X)$. However, (*) characterizes the distribution of T because: (i) $\mathcal{O}_n$ is compact and (ii) $\mathcal{O}_n$ acts transitively on the set of $n \times n$ rank p orthogonal projections (see Nachbin (1967, Chapter 3). The conclusion is that for any function f,

$$\mathcal{L}(f(t(X))|P_0) = \mathcal{L}(f(t(X))|P)$$

as long as $\mathcal{L}(X) = \mathcal{L}(\Gamma X)$ when $\mathcal{L}(X) = P$. These ideas are developed further in Section 4.

Example 1.1 contains the elements of the argument which led to general theorems in both Sections 2 and 3. Once these results are established, the rest of the paper consists of examples from multivariate analysis. In Section 4, we discuss the MANOVA problem in the manner of Example 1.1. In Section 5, we provide sufficient conditions that the sample canonical correlations have the same distribution as if the variables were independent normals. Also, a robustness property of tests based on the multiple correlation coefficient is proved. Some applications to problems involving the complex normal distribution are given in Section 6.

2. MAIN RESULT

In this section, we give our main results together with an elementary application. To set notation, R^k will denote the k-dimensional Euclidean space of column vectors, x' denotes the transpose of $x \in R^k$, $\mathcal{O}_k$ is the group of $k \times k$ orthogonal matrices and $G\ell_k$ is the group of $k \times k$ non-singular matrices. If X is a random vector, $L(X)$ will denote the distribution of X. If $t(X)$ is any statistic, $L(t(X)|P)$ will denote the distribution of $t(X)$ when $L(X) = P$. Also, $N(\mu, I_n \otimes \Sigma)$ denotes the normal distribution on the vector space of $n \times p$ matrices. Here, μ is the mean matrix and $I_n \otimes \Sigma$ is the Kronecker product of the $n \times n$ identity matrix with the $p \times p$ positive definite matrix Σ.

Suppose that $(\mathcal{X},\mathcal{B})$ is a measurable space and G_0 is a group which acts measurably on the left of $\mathcal{X}$. Let $M(\mathcal{X})$ be the set of all probability measures on $(\mathcal{X},\mathcal{B})$. If $P \in M(\mathcal{X})$ and $g \in G_0$, then gP denotes the probability defined by

$$(gP)(B) = P(g^{-1}B),\ B \in \mathcal{B}.$$

The basic situation to be considered here is the following: A measurable space $(\mathcal{X},\mathcal{B})$ is acted on measurably and transitively by a locally compact topological group G. It is assumed that the group G can be represented as $G = K \cdot H$ (each g can be written $g = kh$ for $k \in K$, $h \in H$) where

(i) H is a normal subgroup of G

(ii) K is a compact subgroup of G.

Remark: In most applications K will be the quotient group G/H, but that assumption is not necessary here.

Let μ denote the unique invariant probability measure on K. We assume that the mapping $(k,x) \to kx$ from $K \times \mathcal{X}$ into $\mathcal{X}$ is a jointly measurable mapping. Also, let

$$\mathcal{P}_K = \left\{ P \,\middle|\, \begin{array}{l} P \text{ is a probability on } (\mathcal{X},\mathcal{B}), \\ kP = P \text{ for all } k \in K \end{array} \right\}$$

<u>Theorem 2.1</u>: Suppose that t is a measurable mapping from $(\mathcal{X},\mathcal{B})$ to $(\mathcal{Y},\mathcal{C})$ such that t is H-invariant. Then,

$$\mathcal{L}(t(X)|P) = \mathcal{L}(t(X)|P') \quad \text{for } P,P' \in \mathcal{P}_K.$$

<u>Proof</u>: It suffices to show that for each bounded measurable real valued function f we have

$$\int f(t(x))P(dx) = \int f(t(x))P'(dx) \quad \text{for } P,P' \in \mathcal{P}_K. \tag{2.1}$$

Since $P \in \mathcal{P}_K$, for $k \in K$ it follows that

$$\int f(t(x))P(dx) = \int f(t(kx))P(dx). \tag{2.2}$$

Integrating both sides of (2.2) over K yields

$$\int f(t(x))P(dx) = \int_K \int_{\mathcal{X}} f(t(kx))P(dx)\mu(dk). \tag{2.3}$$

Fix $x_0 \in \mathcal{X}$. Given $x \in \mathcal{X}$, there exists $g \in G$ such that $x = gx_0$ since G acts transitively on $\mathcal{X}$. By assumption $g = k_1h$ for some $k_1 \in K$ and $h \in H$. Thus,

$$(2.4)\qquad f(t(kx)) = f(t(kk_1hx_0)) = f(t(kk_1h(kk_1)^{-1}kk_1x_0))$$
$$= f(t(kk_1x_0))$$

where the last equality follows from the normality of H and the assumed invariance of t. From (2.4) and the invariance of the measure μ, we have

$$(2.5)\qquad \int_K f(t(kx))\mu(dk) = \int_K f(t(kk_1x_0))\mu(dk) = \int_K f(t(kx_0))\mu(dk).$$

Using Fubini's Theorem and substituting (2.5) into (2.3) yields

$$(2.6)\qquad \int_X f(t(x))P(dx) = \int_X \int_K f(t(kx_0))\mu(dk)P(dx)$$
$$= \int_K f(t(kx_0))\mu(dk).$$

Since (2.6) holds for each $P \in \mathcal{P}_K$, (2.1) holds and the proof is complete.

Here is a simple example.

<u>Example 2.1</u>: The canonical form of the univariate ANOVA model is often written as

$$X = \begin{pmatrix} U \\ V \\ W \end{pmatrix} = \begin{pmatrix} \beta \\ \gamma \\ 0 \end{pmatrix} + \epsilon$$

where $U \in R^p$, $V \in R^q$, $W \in R^r$, $\beta \in R^p$, $\gamma \in R^q$, and ϵ is a vector of errors which has a normal distribution with mean zero and covariance $\sigma^2 I_n$ $(n = p + q + r)$. The standard test of the null hypothesis $\gamma = 0$ versus the alternative $\gamma \neq 0$ rejects for large values of

$$t(X) = \frac{\|V\|^2}{\|V\|^2 + \|W\|^2}$$

The sample space $\mathcal{X}$ for this problem is taken to be those $x \in R^n$ such that $\|v\|^2 + \|w\|^2 > 0$ where $x' = (u',v',w')$. In this example, the group G has a typical element (a,b,Γ) where $a \in (0,\infty)$, $b \in R^p$, $\Gamma \in \mathcal{O}_{q+r}$ and the action of G on $\mathcal{X}$ is

$$(a,b,\Gamma)\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} au + b \\ a\Gamma\begin{pmatrix} v \\ w \end{pmatrix} \end{pmatrix}.$$

The composition in G is

$$(a_1,b_1,\Gamma_1)(a_2,b_2,\Gamma_2) = (a_1a_2, a_1b_2+b_1, \Gamma_1\Gamma_2),$$

and G is obviously transitive on $\mathcal{X}$. Take H and K to be

$$H = \{(a,b,\Gamma) | (a,b,\Gamma) \in G,\ \Gamma = I_{q+r}\}$$

$$K = \{(a,b,\Gamma) | (a,b,\Gamma) \in G,\ a = 1,\ b = 0\}.$$

It is easily checked that $G = K \cdot H$, H is normal in G and K is compact. Since t is H-invariant, it follows from Theorem 1 that

$$\mathcal{L}(t(X)|P) = \mathcal{L}(t(X)|P_o)$$

for all $P \in \mathcal{P}_K$ where P_o is the $N(0,I_n)$ distribution. Of course, $\mathcal{L}(t(X)|P_o)$ is a Beta distribution with parameters, q/2 and r/2. A description of $\mathcal{P}_K$ will be given after we establish a general representation result for elements of $\mathcal{P}_K$.

In some situations (see Section 5), the subgroup H of interest is not normal in G but the compact subgroup K is normal in G. In this case, we still have the conclusion of Theorem 2.1. More precisely, assume that

G acts transitively on $\mathcal{X}$ and that $G = K\,H$ where

(i) K is normal in G

(ii) K is a compact subgroup of G.

<u>Theorem 2.2</u>: Suppose that the measurable map t from $(\mathcal{X},\mathcal{B})$ to $(\mathcal{Y},\mathcal{C})$ is invariant under H. Then

$$\mathcal{L}(t(X)|P) = \mathcal{L}(t(X)|P') \quad \text{for } P,P' \in \mathcal{P}_K.$$

<u>Proof</u>: The proof is similar to that of Theorem 2.1. It suffices to establish (2.1). The argument used to prove Theorem 2.1 shows that a given x can be written as k_1hx_0 where x_0 is fixed in $\mathcal{X}$. Thus, for any bounded measurable f on $\mathcal{Y}$, we have

$$f(t(kx)) = f(t(kk_1hx_0)) = f(t(h^{-1}kk_1hx_0)) \tag{2.7}$$

since t is H-invariant. Since equation (2.3) is valid for the case at hand, Fubini's Theorem and (2.7) yield

$$\begin{aligned} \int f(t(x))P(dx) &= \int_K \int_{\mathcal{X}} f(t(kx))P(dx)\mu(dk) \\ &= \int_{\mathcal{X}} \int_K f(t(h^{-1}kk_1hx_0))\mu(dk)P(dx). \end{aligned} \tag{2.8}$$

However, the normality of K on G and the invariance of the probability measure μ on K implies that

$$\int_K f(t(h^{-1}kk_1hx_0))\mu(dk) = \int_K f(t(kx_0))\mu(dk).$$

Combining this with (2.8) yields

$$\int_{\mathcal{X}} f(t(x))P(dx) = \int_K f(t(kx_0))\mu(dk).$$

Thus equation (2.1) holds and the proof is complete.

3. A REPRESENTATION THEOREM

In this section, we give a result from Eaton (1979) which describes elements of the set $\mathcal{P}_K$ occuring in Theorems 2.1 and 2.2. To be precise, assume $(\mathcal{X},\mathcal{B})$ is a measurable space, K is a compact group which acts on $(\mathcal{X},\mathcal{B})$ in such a way that the mapping $(k,x) \to kx$ is measurable from $K \times \mathcal{X}$ to $\mathcal{X}$. The invariant probability measure on K is denoted by μ. Also assume

(3.1) $\mathcal{S}$ is a measurable subset of $(\mathcal{X},\mathcal{B})$ such that $\mathcal{S} \cap \{kx \mid k \in K\}$ consists of exactly one point -- say $s(x)$ -- and the function $x \to s(x)$ from $(\mathcal{X},\mathcal{B})$ to $(\mathcal{S},\mathcal{B}_0)$ is measurable. Here $\mathcal{B}_0$ is the σ-algebra on $\mathcal{S}$ inherited from $(\mathcal{X},\mathcal{B})$.

If R is a probability measure on $(\mathcal{S},\mathcal{B}_0)$, the extension of R to $(\mathcal{X},\mathcal{B})$ will be denoted by $\bar{R}$ -- that is,

$$\bar{R}(B) \equiv R(B \cap \mathcal{S}), \quad B \in \mathcal{B}.$$

As in Section 2,

$$\mathcal{P}_K = \{P \in \mathcal{M}(\mathcal{X}) \mid kP = P,\ k \in K\}.$$

<u>Theorem 3.1</u> (Eaton (1979)): Given the above assumptions, the following are equivalent

(i) $P \in \mathcal{P}_K$

(ii) $P = \int_K k\bar{R}\mu(dk)$ for some $R \in \mathcal{M}(\mathcal{S})$.

<u>Remark (3.1)</u>: Equation (ii) means for each $B \in \mathcal{B}$,

$$P(B) = \int_K (k\bar{R})(B)\mu(dk) = \int_K \bar{R}(k^{-1}B)\mu(dk),$$

or equivalently

$$(3.2) \qquad \int f(x)P(dx) = \int\int f(kx)\bar{R}(dx)\mu(dk)$$

for each bounded measurable f.

<u>Proof</u>: If (ii) holds and $k_1 \in K$, then

$$(k_1P)(B) = P(k_1^{-1}B) = \int_K \bar{R}(k^{-1}k_1^{-1}B)\mu(dk)$$

$$= \int_K \bar{R}(k^{-1}B)\mu(dk) = P(B)$$

so $P \in \mathcal{P}_K$. The third equality follows from the invariance of μ. Conversely, if $P \in \mathcal{P}_K$ we have

$$(3.3) \qquad \int f(x)P(dx) = \int f(kx)P(dx)$$

for all $k \in K$.

Define R on $(\mathcal{S},\mathcal{B}_0)$ by

$$R(B) = P(s^{-1}(B)), \quad B \in \mathcal{B}_0$$

so for each bounded measurable f_1 on $(\mathcal{S},\mathcal{B}_0)$, we have

$$(3.4) \qquad \int_{\mathcal{S}} f_1(s)R(ds) = \int_{\mathcal{X}} f_1(s(x))P(dx).$$

Integrating both sides of (3.3) over K yields

$$(3.5) \qquad \int_{\mathcal{X}} f(x)P(dx) = \int_{\mathcal{X}} [\int_K f(kx)\mu(dk)]P(dx).$$

Given x, there exists a k_1 such that $x = k_1s(x)$. The invariance of μ implies that

$$(3.6) \qquad \int_K f(kx)\mu(dk) = \int_K f(ks(x))\mu(dk).$$

Substituting this into (3.5) and applying (3.4) yields

$$(3.7) \qquad \int_X f(x)P(dx) = \int_X[\int_K f(ks(x))\mu(dk)]P(dx)$$

$$= \int_X[\int_K f(ks)\mu(dk)]R(ds).$$

But, for any bounded measurable function f_2 on X, we have

$$\int_S f_2(s)R(ds) = \int_X f_2(x)\bar{R}(dx)$$

by definition of $\bar{R}$. Applying this to the last member of (3.7) yields

$$(3.8) \qquad \int_X f(x)P(dx) = \int_X \int_K f(kx)\mu(dk)\bar{R}(dx).$$

This is equation (3.2) which is equivalent to (ii) and the proof is complete.

Remark 3.2: The random variable version of Theorem 3.1 goes as follows. Suppose $\mathcal{L}(X) = P$. Then $P \in \mathcal{P}_K$ iff there are independent random variables $U \in K$ and $S \in \mathcal{S}$ such that

(i) $\mathcal{L}(U) = \mu$

(ii) $\mathcal{L}(X) = \mathcal{L}(U(S))$

where $U(S)$ means the random group element U acting on $S \in X$. With this statement of Theorem 3.1, it is obvious that $x \to s(x)$ defines a sufficient statistic for $\mathcal{P}_K$.

Particular cases of Theorem 3.1 are well known in the literature. Fisher (1925) undoubtedly knew this fact when $X = R^n$ and $K = \mathcal{O}_n$. Dawid

(1977) used the random variable version of Theorem 3.1 in his discussion of multivariate linear models. In his application, X is the space $n \times p$ matrices of rank p, $K = O_n$ and there are a couple of natural choices for S (see Section 4).

Example 3.1 (Example 2.1 continued): In the notation of Example 2.1, X is the set of vectors $x \in R^n$ with

$$x = \begin{pmatrix} u \\ v \\ w \end{pmatrix}, \; u \in R^p, \; v \in R^q, \; w \in R^r, \; p + q + n = n$$

where $\|v\|^2 + \|w\|^2 > 0$. Also, K is the compact group whose elements are $(1,0,\Gamma)$ where $\Gamma \in O_{q+r}$ and the group action on X is

$$(1,0,\Gamma)x = \begin{pmatrix} u \\ \Gamma \begin{pmatrix} v \\ w \end{pmatrix} \end{pmatrix} .$$

Let Z be those vectors $z \in R^{q+r}$ with $\|z\| > 0$ and fix $z_0 \in Z$, $\|z_0\| = 1$. Define s on X to X by

$$s(x) = \begin{pmatrix} u \\ \|z\| z_0 \end{pmatrix}, \; x = \begin{pmatrix} u \\ z \end{pmatrix} \in X$$

and let S be the range of s. It is easy to show assumption (3.1) holds, so Theorem 3.1 applies. In terms of random variables, consider $Y \in R^p \times Z$ and partition Y as $Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$, $Y_1 \in R^p$, $Y_2 \in Z$. Then $L(Y) \in P_K$ iff

Y and $\begin{pmatrix} Y_1 \\ \Gamma\, Y_2 \end{pmatrix}$ have the same distribution for all $\Gamma \in \mathcal{O}_{q+r}$. Equivalently,

$L(Y) \in P_K$ iff Y has the representation

$$L(Y) = L\begin{pmatrix} Y_1 \\ RUz_0 \end{pmatrix}$$

where R is a positive random variable, U is uniform on $\mathcal{O}_{q+r}$ and is independent of

$$\begin{pmatrix} Y_1 \\ Rz_0 \end{pmatrix} \in S.$$

4. FIRST APPLICATIONS

In this section we apply the techniques described in the previous sections to two classical problems in multivariate analysis -- namely, the MANOVA problem and the problem of testing for the equality of two covariance matrices. A canonical form of the MANOVA model can be written

$$X = \begin{pmatrix} U \\ V \\ W \end{pmatrix} = \begin{pmatrix} B_1 \\ B_2 \\ 0 \end{pmatrix} + E$$

where X is $n \times p$, U and B_1 are $n_1 \times p$, V and B_i are $n_2 \times p$, W is $n_3 \times p$, and E, the matrix of errors, is $n \times p$. We assume $n_3 \geq p$. The MANOVA problem is to test $H_0 : B_2 = 0$ versus $H_1 : B_2 \neq 0$. The sample space for this example is $\mathcal{X}$ -- the space of $(n_1+n_2+n_3) \times p$ matrices

$$x = \begin{pmatrix} u \\ z \end{pmatrix} = \begin{pmatrix} u \\ v \\ w \end{pmatrix} \quad , \; u : n_1 \times p, \; z : (n_2+n_3) \times p$$

such that z has rank p. When E is $N(0, I_n \otimes \Sigma)$ with Σ positive definite and unknown, a standard invariance argument (see Eaton (1972), Chapter 9) shows that all fully invariant tests are based on $t_0(X)$ which is the vector of the ordered non-zero eigenvalues of

$$V(V'V + W'W)^{-1}V' = V(Z'Z)^{-1}V'$$

where

$$Z = \begin{pmatrix} V \\ W \end{pmatrix} .$$

Let P_o denote the $N(0, I_n \otimes I_p)$ distribution on $\mathcal{X}$. It is well known that $\mathcal{L}(t_o(X)|P_o)$ is the same as the distribution of $t_o(X)$ when

$$\mathcal{L}(X) = N\left[\begin{pmatrix} B_1 \\ 0 \\ 0 \end{pmatrix}, I_n \otimes \Sigma\right]$$

for any B_1 and any Σ.

We will now apply Theorem 2.1 to obtain a larger class of distributions for which the distribution of $t_o(X)$ is the same as when $\mathcal{L}(X) = P_o$. The technique in Dawid (1977) will also yield our results for this example. To apply Theorem 2.1, consider the group G whose elements are (Γ, A, C) where $\Gamma \in \mathcal{O}_{n_2+n_3}$, $A \in G\ell p$ and C is an $n_1 \times p$ matrix. The action of G on $\mathcal{X}$ is

$$(\Gamma, A, C)\begin{pmatrix} u \\ z \end{pmatrix} = \begin{pmatrix} uA'+C \\ \Gamma z A' \end{pmatrix},$$

and the group operation is

$$(\Gamma_1, A_1, C_1)(\Gamma_2, A_2, C_2) = (\Gamma_1\Gamma_2, A_1A_2, C_2A_1' + C_1).$$

The action of G is transitive. With

$$H = \{(\Gamma, A, C) \in G | \Gamma = I_{n_2+n_3}\}$$

and

$$K = \{(\Gamma, A, C) \in G | A = I_p,\ C = 0\}$$

it follows that $G = K \cdot H$, H is normal in G and K is compact.

Let $\mathcal{Y}$ be the set of all $(n_2+n_3) \times (n_2+n_3)$ rank p orthogonal projections on $R^{n_2+n_3}$ and equip $\mathcal{Y}$ with the usual topology. The function t on $\mathcal{X}$ to $\mathcal{Y}$ defined by

$$t(x) = z(z'z)^{-1}z', \quad x = \begin{pmatrix} u \\ z \end{pmatrix}$$

is measurable and is H invariant. Note that t_0 defined earlier is a function of t since the upper left $n_2 \times n_2$ block of t(x) is $v(z'z)^{-1}v' = v(v'v + w'w')^{-1}v'$. By Theorem 2.1, $\mathcal{L}(t(X)|P) = \mathcal{L}(t(X)|P_0)$ for any $P \in \mathcal{P}_K$. In particular, if the distribution of X satisfies

$$(4.1) \qquad \mathcal{L}(X) = \mathcal{L}\left(\begin{pmatrix} U \\ Z \end{pmatrix}\right) = \mathcal{L}\left(\begin{pmatrix} U \\ \Gamma Z \end{pmatrix}\right)$$

for $\Gamma \in \mathcal{O}_{n_2+n_3}$, then the distribution of t(X) is the same as if X is $N(0, I_n \otimes I_p)$. Since t_0 is a function of t, the same conclusion holds for t_0. The results of Section 3 can be used to represent distributions satisfying (4.1). If U is not present in (4.1), this representation is that given in Dawid (1977).

We now turn to a brief discussion of testing for the equality of two covariance matrices. For simplicity, the case of zero means is treated -- the general case can be handled by a similar argument. For this problem, consider two independent data matrices $X_i : n_i \times p$, $i = 1,2$. When $\mathcal{L}(X_i) = N(0, I_{n_i} \otimes \Sigma_i)$ and we wish to test $H_0 : \Sigma_1 = \Sigma_2$ versus $H_1 : \Sigma_1 \neq \Sigma_2$, fully invariant tests are based on the p non-zero eigenvalues of $X_1(X_1'X_1 + X_2X_2')^{-1}X_1'$. Set

$$Z = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

and

$$(4.2) \qquad t(Z) = Z'(Z'Z)^{-1}Z'.$$

Proceeding as in the MANOVA problem (with U absent) shows that the distribution of $t(Z)$ when Z is $N(0, I_n \otimes I_p)$ is the same as when $L(Z) = P \in \mathcal{P}_K$ where $K = \mathcal{O}_{n_1+n_2}$ for this problem. Thus the distributions of fully invariant tests will be the same under a $N(0, I_n \otimes I_p)$ distribution as under any distribution for Z which satisfies $L(Z) = L(\Gamma Z)$, $\Gamma \in \mathcal{O}_{n_1+n_2}$.

The representation of the $\mathcal{O}_{n_1+n_2}$ invariant distributions provided by Theorem 3.1 and Remark 3.2 is the same as that given by Dawid (1977). To see this, let $\mathcal{X}$ be those $x : (n_1+n_2) \times p$ with rank p (so $p \le n_1+n_2$). Pick S to be those x's of the form

$$x = \begin{pmatrix} \alpha \\ 0 \end{pmatrix}, \ \alpha \in G_U^+$$

where G_U^+ is the group of $p \times p$ upper triangular matrices with positive diagonal elements. Assumption (3.1) is verified by using the fact that each $x \in \mathcal{X}$ can be uniquely written as $x = \psi\alpha$ where $\alpha \in G_U^+$ and $\psi : (n_1+n_2) \times p$ satisfies $\psi'\psi = I_p$. From Remark 3.2, $L(Z) \in \mathcal{P}_K$ iff Z has the representation

$$L(Z) = L(U \begin{pmatrix} \alpha \\ 0 \end{pmatrix})$$

where U is uniform on $\mathcal{O}_{n_1+n_2}$ and is independent of $\alpha \in G_U^+$. The distribution of α is arbitrary. In certain applications, it is more convenient to make a different choice for S -- namely pick S to be those x's of the form

$$x = \begin{pmatrix} \beta \\ 0 \end{pmatrix}$$

where β is $p \times p$ and positive definite. Then $\mathcal{L}(Z) \in \mathcal{P}_K$ iff Z has the representation

$$(Z) = (U \begin{pmatrix} \beta \\ 0 \end{pmatrix})$$

where U is uniform on $\mathcal{O}_{n_1+n_2}$ and is independent of β. The distribution of β on $p \times p$ positive definites is arbitrary.

<u>Remark 4.1</u>: When Z is a random $n \times p$ matrix of rank p and if $\mathcal{L}(\Gamma Z) = \mathcal{L}(Z)$, $\Gamma \in \mathcal{O}_n$, then the distribution of

$$Q = Q(Z)Q \equiv Z(Z'Z)^{-1}Z$$

can be characterized as follows. First, Q takes values in the set of $n \times n$ rank p orthogonal projections -- say $\mathcal{Y}_{n,p}$. Since

$$Q(\Gamma Z) = \Gamma Q(Z)\Gamma',$$

and $\mathcal{L}(\Gamma Z) = \mathcal{L}(Z)$, it follows that

$$\mathcal{L}(Q) = \mathcal{L}(\Gamma Q \Gamma'), \ \Gamma \in \mathcal{O}_n.$$

Since the compact group $\mathcal{O}_n$ acts transitively on $\mathcal{Y}_{n,p}$ $(y \to \Gamma y \Gamma')$ there is a unique $\mathcal{O}_n$-invariant probability distribution on $\mathcal{Y}_{n,p}$ which we will call the uniform distribution (see Nachbin (1965), Chapter 3). Now, the distribution of $Q \in \mathcal{Y}_{n,p}$ is invariant so Q must have the uniform distribution on $\mathcal{Y}_{n,p}$. These ideas will appear again in the next section.

<u>Remark 4.2</u>: The Generalized MANOVA problem was introduced in Potthoff and Roy (1964) and discussed at length in Olkin and Gleser (1966) and Kariya

(1978). The techniques used on the MANOVA problem above can also be used on the GMANOVA problem to yield corresponding results. The details are left to the reader.

5. CANONICAL CORRELATIONS

In this section, we discuss the distributions of canonical correlations. Without essential loss of generality we consider the mean zero case. The sample space for this section is $\mathcal{X}$ -- the set of $n \times p$ matrices of rank p. Consider $Z \in \mathcal{X}$ and partition Z as $Z = (Z_1, Z_2)$ where Z_i is $n \times p_i$, $i = 1,2$. The orthogonal projections

$$Q_i = Z_i(Z_i'Z_i)^{-1}Z_i', \quad i = 1,2$$

are elements of $\mathcal{Y}_{n,p_i}$, $i = 1,2$ defined in Section 4 (see Remark 4.1). The <u>canonical correlations</u> are defined to be the $r \equiv \min\{p_1, p_2\}$ largest eigenvalues of Q_1Q_2. To see that this definition agrees with the standard definition in terms of the sample covariance matrix $S = Z'Z$, partition S as

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \; ; \; S_{ij} : p_i \times p_j$$

for $i,j = 1,2$. Classically, the canonical correlations, defined to be the r-largest eigenvalues of $S_{11}^{-1}S_{12}S_{22}^{-1}S_{21}$. But $S_{ij} = Z_i'Z_j$ so

$$Q_1Q_2 = Z_1S_{11}^{-1}S_{12}S_{22}^{-1}Z_2'.$$

However, the non-zero eigenvalues of $Z_1S_{11}^{-1}S_{12}S_{22}^{-1}Z_2'$ (of which there are at most r) are the same as the non-zero eigenvalues of

$$S_{11}^{-1}S_{12}S_{22}^{-1}Z_2'Z_1 = S_{11}^{-1}S_{12}S_{22}^{-1}S_{21}.$$

Thus, our definition coincides with the usual definition.

Given $Z \in X$, let $t(Z)$ be the vector of the r largest eigenvalues (arranged in order) of $Q_1 Q_2$. When Z is $N(0, I_n \otimes I_p) \equiv P_0$, the density of $t(Z)$ is known (see Anderson (1958), Chapter 13). Here we will describe a large class of distributions under which $t(Z)$ has the same distribution as when $L(Z) = P_0$. Consider the group G whose elements are (ψ, Γ, A, B) with $\psi, \Gamma \in O_n$, $A \in G\ell_{p_1}$ and $B \in G\ell_{p_2}$. The action of G on X is

$$(\psi, \Gamma, A, B)(z_1, z_2) = (\psi z_1 A', \Gamma z_2 B')$$

and G acts transitively on X. The group operation is

$$(\psi_1, \Gamma_1, A_1, B_1)(\psi_2, \Gamma_2, A_2, B_2) = (\psi_1\psi_2, \Gamma_1\Gamma_2, A_1A_2, B_1B_2).$$

Let

$$H = \{(\psi, \Gamma, A, B) \in G | \psi = \Gamma\}$$

and

$$K = \{(\psi, \Gamma, A, B) \in G | \Gamma = I_n, A = I_{p_1}, B = I_{p_2}\}.$$

Then $G = K \cdot H$, K is compact and K is normal in G (but H is not normal in G). Thus Theorem 2.2 is applicable and we have that

$$L(t(X) | P_0) = L(t(Z) | P)$$

for all $P \in P_K$ since $P_0 \in P_K$. To describe P_K, first observe that elements of K act on X by

$$(\psi, I_n, I_{p_1}, I_{p_2})(z_1, z_2) = (\psi z_1, z_2).$$

Thus $L(Z) \in P_K$ iff $Z = (Z_1, Z_2)$ has the same distribution as $(\psi Z_1, Z_2)$, $\psi \in O_n$. This certainly occurs when

$$\mathcal{L}(Z) = N(0, I_n \otimes \bar{\Sigma})$$

where

$$\bar{\Sigma} = \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix}, \ \Sigma_{ii} : p_i \times p_i,\ i = 1,2.$$

A wider class of distributions for which $\mathcal{L}(Z) \in \mathcal{P}_K$ can be constructed as follows. Given a function q defined on $[0,\infty)$ to $[0,\infty)$ such that

(5.1) $$\int_{\mathcal{X}} q(\mathrm{tr}\ x'x)dx = 1,$$

let

(5.2) $$f(x|\Sigma) = |\Sigma|^{-n/2}\ q(\mathrm{tr}\ \Sigma^{-1}x'x).$$

Then $f(.|\Sigma)$ is a density on $\mathcal{X}$ for each positive definite $\Sigma : p \times p$. Partition Σ as

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \ \Sigma_{ij} : p_i \times p_j.$$

When $\Sigma_{12} = 0$ and Z has the density $f(.|\Sigma)$ on $\mathcal{X}$, then a routine calculation shows that $\mathcal{L}(Z) \in \mathcal{P}_K$.

We end this section with a robustness property of a test based on the multiple correlation coefficient. Assume $p_1 = 1$ so $p_2 = p-1$ and $r = 1$. In this case Q_1Q_2 has one non-zero eigenvalue and it is R_M^2, where R_M denotes the multiple correlation coefficient. When Z is $N(0, I_n \times \Sigma)$, Σ unknown, the test of $H_0 : \Sigma_{12} = 0$ versus $H_1 : \Sigma_{12} \neq 0$ which rejects for large values of R_M^2 is a uniformly most powerful invariant (UMPI) (under the subgroup H above) test of H_0 versus H_1. Fix q satisfying

(5.1) and assume q is a convex. Let F be those distributions on X which have a density of the form (5.2) and let $F_0 \subseteq F$ be those distributions in F with $\Sigma_{12} = 0$. For testing $H_0 : L(Z) \in F_0$ versus $H_1 : L(Z) \in F - F_0$, we will show that rejecting for large values of R_M^2 is a UMPI test. First, this testing problem is invariant under the action of the group H on the sample X. This action induces a group action on the parameter space of Σ's in the obvious way -- namely, if

$$h = (\Gamma, \Gamma, A, B) \in H,$$

then

$$\bar{h}(\Sigma) = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \Sigma \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}'$$

so if $f(.|\Sigma)$ is the density of Z, then $f(.|\bar{h}(\Sigma))$ is the density of $h(Z)$. A routine calculation shows that a maximal invariant in the sample space is

$$t(Z) \equiv R_M^2 = (Z_1'Z_1)^{-1} Z_1'Z_2(Z_2'Z_2)^{-1} Z_2'Z_1 \in [0,1]$$

and a maximal invariant in the parameter space is

$$\delta = \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \in [0,1].$$

Now, the argument parallels that in Kariya (1981a) regarding Hotelling's T^2.

<u>Lemma 5.1</u>: Let P_δ be the distribution of R_M^2 when the parameter value is δ. Then the Radon-Nikodym derivative dP_δ/dP_0 is

$$(5.3) \qquad h(t(Z)|\delta) = \xi(\delta)/\xi(0)$$

where

$$(5.4) \qquad \xi(\delta) = \int_{G\ell_1} \int_{G\ell_{p-1}} q(a^2+2ab_{11}t^{\frac{1}{2}}\delta^{\frac{1}{2}}+\mathrm{tr}BB')|a|^{n-1}|BB'|^{(n-p+1)/2}dadB.$$

Here, b_{11} is the (1,1) element of $B \epsilon G\ell_{p-1}$.

<u>Proof</u>: As in Lemma 3.1 of Kariya (1981a), we apply Wijsman's Theorem (1967) to obtain $h(.|\delta)$ is N_δ/N_0 where

$$(5.5) \qquad N_\delta = \int_{G\ell_1 \times G\ell_{p-1}} q(\mathrm{tr}C'Z'ZC\Sigma^{-1})|C'C|^{n/2}\, v_1(da)v_2(dB).$$

In this formula,

$$C = \begin{pmatrix} a & 0 \\ 0 & B \end{pmatrix} \epsilon\ G\ell_1\ G\ell_{p-1}$$

and

$$v_1(da) = |a|^{-1}da, \quad v_2(dB) = |B'B|^{-(p-1)/2}dB$$

are left and right invariant measures on G_1 and G_{p-1} respectively. To show $N_\delta/N_0 = \xi(\delta)/\xi(0)$, we need to make a change of variables in (5.5). First, there exists an element $h_1 \epsilon H$ such that

$$Z'Z = \bar{h}_1 \left[\begin{pmatrix} 1 & t^{\frac{1}{2}}u_0' \\ t^{\frac{1}{2}}u_0 & I_{p-1} \end{pmatrix}\right] \equiv \bar{h}_1(S(t)),\ t \epsilon [0,1],$$

where $u_0' = (1,0,\ldots,0)$, $u_0 \epsilon R^{p-1}$. Also, there exists an element $h_2 \epsilon H$ such that

$$\Sigma^{-1} = \bar{h}_2 \left[\begin{pmatrix} 1 & \lambda u_0' \\ \lambda u_0 & I_{p-1} \end{pmatrix}\right] \equiv \bar{h}_2(\Lambda(\lambda)), \quad \lambda \in [0,1].$$

With

$$h_i = (\Gamma_i, \Gamma_i, a_i, B_i), \; i = 1,2,.$$

and

$$C_i = \begin{pmatrix} a_i & 0 \\ 0 & B_i \end{pmatrix}, \; i = 1,2$$

we have

$$Z'Z = C_1 S(t) C_1'$$

and

$$\Sigma^{-1} = C_2 \Lambda(\lambda) C_2'.$$

Replacing C by $C_1' C C_2$ in (5.5), the invariance of v_1 and v_2 show that N_δ/N_o is $\bar{N}_\delta/\bar{N}_{(0)}$ where

(5.6) $$\bar{N}_\delta = \int_{G\ell_1 \times G\ell_{p-1}} q(\mathrm{tr}\, C'S(t)C\Lambda(\lambda))|C'C|^{n/2} v_1(da) v_2(dB).$$

Noting that $\lambda = \delta^{\frac{1}{2}}$ and expanding the trace in (5.5) yields

$$\mathrm{tr}\, C'S(t)C\Lambda(\lambda) = a^2 + 2ab_{11} t^{\frac{1}{2}} \delta^{\frac{1}{2}} + \mathrm{tr}\, BB'.$$

Substituting this into (5.6) shows that $\bar{N}_\delta/\bar{N}_0 = \xi(\delta)/\xi(0)$ which completes the proof.

Now, let $\varphi(u)$ denote the right hand side of (5.4) with $t^{\frac{1}{2}}$ replaced by u. Then changing a to -a in (5.4) shows that $\varphi(u) = \varphi(-u)$. The convexity of q implies that for $\frac{1}{2} < \alpha < 1$,

$$\varphi(u) = \varphi(u) + (1-\alpha)\varphi(-u) \geq \varphi((2\alpha-1)u)$$

so φ is non-decreasing on [0,1].

The Neyman-Pearson Lemma and our previous results yield:

Theorem 5.1: For testing H_0 versus H_1, the test which rejects for large values of R_M^2 is UMPI. Under H_0, the distribution of R_M^2 is the same as when $\mathcal{L}(Z) = N(0, I_n \otimes I_p)$.

Remark 5.1: This result is also valid for the case of non-zero means. The argument is a minor variation of that given above (also, see Kariya (1981a)).

6. SOME COMPLEX NORMAL PROBLEMS

In this section we discuss two problems related to some recent results of Andersson and Perlman (1981). To describe the situation, suppose that we have a random sample with

$$L\left[\begin{pmatrix} X_i \\ Y_i \end{pmatrix}\right] = N(0,\Sigma),\ i = 1,\ldots,n$$

with $X_i \in R^p$ and $Y_i \in R^p$ and partition Σ as

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},\ \Sigma_{ij} : p \times p.$$

Consider the following three classes of $(2p) \times (2p)$ covariances:

$$C_1 = \{\Sigma | \Sigma \text{ is positive definite}\}$$

$$C_2 = \{\Sigma | \Sigma \in C_1,\ \Sigma_{11} = \Sigma_{22},\ \Sigma_{12} = -\Sigma_{21}\}$$

$$C_3 = \{\Sigma | \Sigma \in C_1,\ \Sigma_{11} = \Sigma_{22},\ \Sigma_{12} = \Sigma_{21} = 0\}.$$

Khatri (1965) considered the problem of testing $H_0^{(1)} : \Sigma \in C_3$ versus $H_1^{(1)} : \Sigma \in C_2$. Elements of C_2 are usually said to have "complex structure" while those in C_1 and C_3 are said to have "real structure" -- see Goodman (1963) and Brillinger (1974). The above testing problem can be interpreted as testing that a complex normal random vector is in fact real. In contrast, the problem of testing $H_0^{(2)} : \Sigma \in C_2$ versus $H_1^{(2)} : \Sigma \in C_1$ is testing that a real normal is in fact complex. Both of these problems are discussed in detail in Andersson and Perlman (1981). They reduce both problems via invariance and establish many properties of invariant tests. In what follows, we apply the results of Section 2 to show that the null distribution of all invariant tests is the same under normality as under a wider class of distributions. The result and techniques are similar to those in Sections 4 and 5.

First, we treat testing $H_0^{(1)}$ versus $H_1^{(1)}$. Write the data in matrix form to yield $W : (2n) \times p$ whose first n rows are $X_1', \ldots, X_n'$ and whose second n rows are $Y_1', \ldots, Y_n'$. Under $H_0^{(1)}$,

$$\mathcal{L}(W) = N(0, I_{2n} \otimes \Sigma_{11}).$$

The sample space for W is taken to be $\mathcal{X}$ -- the set of real $(2n) \times p$ matrices of rank p. Take G to be the group whose elements are (Γ, A) with $\Gamma \in \mathcal{O}_{2n}$, $A \in G\ell_p$, and group operation $(\Gamma_1, A_1)(\Gamma_2, A_2) = (\Gamma_1\Gamma_2, A_1A_2)$. The action of G on $\mathcal{X}$ is $(\Gamma, A)W = \Gamma W A'$, so G is transitive on $\mathcal{X}$. Also, take

$$H = \{(\Gamma, A) \in G | \Gamma = I_{2n}\}$$

and

$$K = \{(\Gamma, A) | A = I_p\}$$

so K is compact and both H and K are normal in G. The function

$$t(W) = W(W'W)^{-1}W'$$

is a maximal invariant under H.

Theorem 2.1 implies that the distribution of t(W) under a $N(0, I_{2n} \otimes I_p)$ distribution for W is the same as under any distribution for W which satisfies $\mathcal{L}(W) = \mathcal{L}(\Gamma W)$ for $\Gamma \in \mathcal{O}_{2n}$. Now, all the tests of $H_0^{(1)}$ versus $H_1^{(1)}$ discussed in Andersson and Perlman (1981) are invariant under the group H and are thus functions of $t(W)$. Hence the null distribution of all these tests is the same as when $L(W) = L(\Gamma W)$ for $\Gamma \epsilon O_{2n}$.

To discuss testing $H_0^{(2)}$ and $H_1^{(2)}$, it is convenient to introduce the complex random vectors

$$Z_j = X_j + iY_j, \; j = 1,\ldots,n$$

and form the data matrix $Z : n \times p$ with rows $Z_1^*,\ldots,Z_n^*$ where * denotes conjugate transpose. Under $H_0^{(2)}$, Z has a complex normal distribution,

$$L(Z) = ¢N(0, I_n \otimes H)$$

where $H = \Sigma_{11} + i\Sigma_{12}$ and

$$\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \epsilon \; C_2.$$

Of course H is Hermitian and positive definite. The sample space for Z is taken to be X -- the space of all $n \times p$ complex matrices of rank p. To apply Theorem 2.1, consider the group G whose elements are (U,A) where U is an element of U_n (the group of $n \times n$ unitary matrices) and $A \epsilon ¢G\ell_p$ (the group of $p \times p$ non-singular complex matrices). The action of G on X is

$$(U,A)Z = UZA^*$$

so G is transitive on X. Also take

$$H = \{(U,A) \in G | U = I_n\}$$

and

$$K = |(U,A) \in G | A = I_p\}$$

so K is compact and both are normal in G.

A maximal invariant under the action of H on is

$$t(Z) = Z(Z^*Z)^{-1}Z^*.$$

Let P_o denote the $\mathbb{C}N(0,I_n \otimes I_p)$. Theorem 2.1 implies that $L(t(Z)|P_o) = L(t(Z)|P)$ for any probability measure P for which $L(Z) = L(UZ)$, $U \in \mathcal{U}_n$. All of the tests discussed in Andersson and Perlman (1981) are H invariant and thus functions of t(Z). Hence the null distribution under P_o is the same as the null distribution under any P for which $L(Z) = L(UZ)$, $U \in \mathcal{U}_n$.

REFERENCES

[1] Anderson, T.W. (1958). An Introduction to Multivariate Statistical Analysis. Wiley, New York.

[2] Andersson, S.A. and Perlman, M.D. (1979). Two testing problems relating the real and complex multivariate normal distributions. Technical Report #95, Department of Statistics, University of Chicago.

[3] Brillinger, D.R. (1974). Time Series: Data Analysis and Theory. Holt, Rinehart, and Winston, New York.

[4] Dawid, A.P. (1977). Spherical matrix distributions and a multivariate model. J. Roy. Statist. Soc., Ser. B, 39, 254-261.

[5] Eaton, M.L. (1972). Multivariate Analysis. University of Copenhagen.

[6] Eaton, M.L. (1979). Unpublished manuscript.

[7] Efron, B. (1969). Student's t-test under symmetry conditions. J. Amer. Statist. Assn., 64, 1278-1302.

[8] Fraser, D.A.S. and Ng, K.W. (1980). Multivariate regression analysis with spherical error. Multivariate Analysis V (ed. by P.R. Krishnaiah). 369-386. North Holland, New York.

[9] Gleser, L. and Olkin, I. (1966). A k-sample regression model with covariance. Multivariate Analysis (ed. P.R. Krishnaiah), 59-72. Academic Press, New York.

[10] Goodman, N.R. (1963). Statistical analysis based on a certain multivariate complex Gaussian distribution(an introduction). Ann. Math. Statist., 34, 152-176.

[11] Kariya, T. (1978). The general MANOVA problem. Ann. Statist., 6, 200-213.

[12] Kariya, T. (1981a). A robustness property of Hotelling's T^2-test. Ann. Statist., 9, 210-214.

[13] Kariya, T. (1981b). The uniqueness of null distributions of multivariate tests in the class of left O_n-invariant distributions and a robustness property of the MANOVA tests. Technical report, University of Pittsburgh.

[14] Kariya, T. and Eaton, M.L. (1978). Robust tests for spherical symmetry. Ann. Statist., 5, 206-215.

[15] Khatri, C.G. (1965). A test for reality of a covariance matrix in a certain complex Gaussian distribution. Ann. Math. Statist., 36, 115-119.

[16] Nachbin, L. (1965). The Haar Integral. Van Nostrand, New York.

[17] Potthoff, R.F. and Roy, S.N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. Biometrika, 51, 313-326.

[18] Wijsman, R.A. (1967). Cross-sections of orbits and their application to densities of maximal invariants. Fifth Berkely Symp. Math. Statist. Prob., 1, 389-400.